Accepted to Interact 2007

An Eye Tracking Study of How Font Size, Font Type, and Pictures Influence Online Reading

David Beymer[1], Daniel M Russell[2], and Peter Z Orton[3]

[1] IBM Almaden Research Center, 650 Harry Road,
San Jose, California 95120 USA
[2] Google, 1600 Amphitheatre Parkway,
Mountain View, California 94043 USA
[3] IBM Center for Advanced Learning, 20 Old Post Road,
Armonk, New York 10504 USA
{beymer, porton}@us.ibm.com, drussell@google.com

Abstract. In order to maximize online reading performance and comprehension, how should a designer choose typographical variables such as font size and font type? Do accompanying pictures distract the reader? This paper presents an eye tracking study of how font size, font type, and pictures affect online reading. In a between-subjects design, we collected data from 114 subjects reading stories formatted in a variety of point sizes, san serif and serif fonts, and containing different types of pictures. Reading statistics such as reading speed and regressions (backward eye movements) were automatically computed using the WebGazeAnalyzer system. While our results for font size and font type yielded only one significant difference in reading, we did find significant differences in speed and regressions when pictures related to the text were replaced with advertisements. Similar significant results were found for demographic variables such as age group and whether English is the subject's first language.

Keywords: Eye tracking, typography, font size, font type, online reading.

1 Introduction

From a design standpoint, how does one adjust the typography of a web page or online document for optimal reading? What guidelines do designers have for how embedded pictures influence the reading process? In this paper, we present studies that address three design issues: (1) what is the best font size for reading online, (2) which font type, serif or san serif, is best for reading, and (3) how pictures distract readers. We are motivated by the design of online e-learning material, so retention of the material is as important a factor as speed and efficiency.

These typographic issues have been studied by advertisers, psychologists, ergonomists, and designers for over 100 years, focusing mostly on paper but now moving to address online presentation on computer screens. For example, in the font size issue, using too small a font makes the letters illegible, while too large a font needlessly wastes page space. Paterson and Tinker, who in the 1920s – 1940s studied a number of typographical issues [1], found that for paper, 10 pt text was read faster than 6, 8, 12, or 14 pt [2]. Measuring character size by visual angle, Legge, *et al* [3] found that reading speed was fairly constant for a range of character sizes, 0.3º-2º, but it deteriorates outside of this range. Looking at font sizes of 10, 12 and 14 pt on a computer screen, Bernard, *et al* [4] found that 12 pt was read the fastest.

For font type, there is a lack of statistically significant differences in font studies to rule in favor of serif or san serif fonts. Those favoring serif fonts claim that the serif brackets and the contrasting use of thin and thick lines makes the letters and words more distinctive and hence easier on the eye. According to those favoring san serif, those same shifts in line width create an exaggerated contrast that impairs reading speed. In a study of 10 fonts types on paper, Paterson and Tinker [5] isolated 2 fonts as poor performers, but those fonts are not in use today (American typewriter, Old English). In a more recent study, Boyarski *et al* [6] tested modern computer fonts, explicitly comparing serif vs. san serif fonts for computer screens. There was a comprehension advantage for Georgia (serif) over Verdana (san serif), but no reading speed difference.

Web pages in the online world generally have much richer combinations of text and pictures than printed material. Given our increased use of pictures, what is the interplay between text and pictures? Do different types of pictures (advertisements, pictures related to the text) affect reading differently? Web design guidelines call for restricted use of images (partially, although, from download time constraints) and to use images related to the main body of text [7]. Web users have learned to filter out ads and meaningless graphics – this is especially true of banner ads, from which the term *banner blindness* has arisen [8]. While Burke, *et al* [9] studied the negative effect of banner ads on a search task, we have not found a study of how pictures influence the reading process.

To understand the detailed structure of how people read text, psychologists and other researchers have turned to eye gaze tracking as a valuable analysis tool. In eye gaze tracking, a camera tracks and records where a subject's eye is looking; these gaze points are mapped to the text to follow the subjects' reading behavior. Early eye gaze tracking practitioners working 100 years ago used crude recording techniques [1], bouncing a beam of light off the subjects' eye and recording the reflection on film. Eye trackers today bounce invisible infrared light off the subjects' eye and track it using digital cameras mounted near the monitor. The output is the same, a stream of *(x, y)* gaze coordinates over time.

Eye tracking analysis has revealed how the eye moves during the reading process – see Rayner and Pollatsek [10] for an excellent summary. The eye reads a line of text in discrete chunks by making a series of fixations and saccades (Fig. 1). A *fixation* is a brief moment, around 250 ms, where the eye is paused on a word or word group, and the brain processes the visual information. A *saccade* is a fast eye movement, usually forward in the text around 8-12 characters, to take in the next section of text. A *regression* is a backwards motion in the text, and it indicates confusion. A *return sweep* is the eye motion from the end of one line of text to the beginning of the next. The *perceptual span* refers to the size of the visual window processed at each fixation.

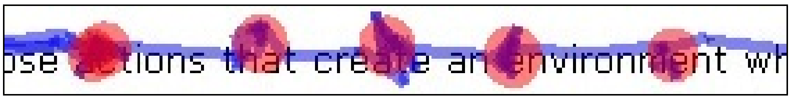


Fig. 1 People read text in discrete chunks of fixations (circles) connected by jumps called saccades. The raw eye gaze is in blue.

Paterson and Tinker used eye gaze tracking to study the typographical issues of font size and font type [11,12]. For font size, they compared 10 pt (the optimal size, according to their previous reading speed study) against 6 pt and 14 pt, and they replicated their previous speed findings. The 6 pt font was slower due to increased fixation duration, which was probably caused by reduced character visibility. For 14 pt, there were more fixations and thus probably a smaller perceptual span. The perceptual span result, however, is contradicted in a later study by Morrison & Rayner [13], who show that saccade length depends on character spacing, not visual angle – that is, saccades should scale up for larger fonts. For font type, Paterson & Tinker compared Scotch Roman with Old English to further examine a large reading speed difference from [5]. Old English was slower to read because of an increased number of fixations and an increased regression rate.

Eye tracking analysis has also been used to investigate the relationship between text and pictures on web pages. In the Eyetrack III study [14], media researchers studied how subjects read online news sites. Among their results, they found (a) ads mixed in with the main test are viewed more than ads in the periphery, (b) size matters for ads, with larger ones viewed more than smaller, and (c) ads with text that blend in with the main page text are viewed more. Also using eye tracking, the Norman Nielsen Group [15] recommends that pictures relate to content and don't look like ads.

In this paper, we use eye tracking to investigate how font size, font type, and pictures influence online reading. All the Paterson and Tinker studies were done on paper – will we see the same results on computer screens, and using modern eye tracking equipment? Will saccade length vary linearly with font size, as suggested by [13]? We will use modern computer fonts for the study of font type, as opposed to Scotch Roman and Old English. And for the issue of pictures, we will test how changing the picture type (related to content vs. ads vs. no picture) affects reading statistics and retention.

To collect and analyze the experimental data, we use WebGazeAnalyzer (WGA) [16, 17], a tool that records eye gaze in the context of a web browsing session. WGA's web browser is instrumented to record all URLs visited and HTML content, so at analysis time, WGA can automatically map eye gaze to web page text and graphics. By designing our experiment as a series of web pages, we can use WGA to compute reading statistics and thus address our issues in typography and picture viewing.

2 Experiment

In our experiment, we examine how font size, font type, and pictures influence online reading. Using a between-subjects design, we use WGA to collect eye tracking-based reading statistics on subject pools assigned different font sizes, font types, etc. We now describe our experimental design and method of data collection.

2.1 Typographic Issues

Table 1 summarizes the conditions tested for each typographic issue. A one-page story was assigned to each typographic issue and formatted appropriately (e.g. small, medium, and large fonts) for each subject group. The stories were taken from a science news web site with universally appealing stories written at an 8th grade reading level; the content was selected to go beyond common knowledge to allow for testing of retention.

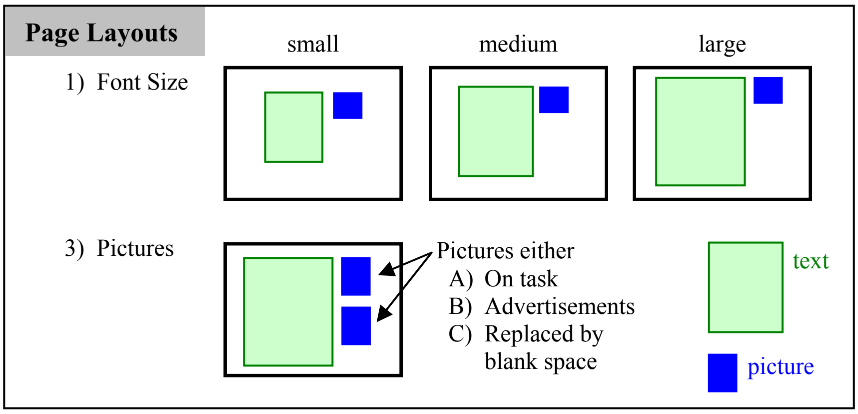
Table 1. Our study of online reading examines the three typographic issues below.

Typographic Issue	Conditions		N	
1) Font Size	A) small	10 pt	28	
	B) medium	12 pt	27	(82 total)
	C) large	14 pt	27	
2) Font Type	A) san serif	Helvetica	41	(82 total)
	B) serif	Georgia	41	
3) Pictures	A) on task	pictures relate to text	27	
	B) advert	pictures are ads, same size as A	27	(82 total)
	C) none	pictures replaced by blank space	28	

Stories are formatted on a single page to avoid the need for scrolling. The text color is black against a white background, and the fonts are antialiased. Within paragraph structure, line breaks are constrained to occur at the same location across formatting conditions, preserving line structure for each story. For example, for font size, the exact same paragraph formatting is used across conditions, resulting in different magnifications of the same page (see Table 2). This avoids confounding font size with paragraph formatting. For the font type task, layout is nearly identical across condition, and in the pictures task, the pictures are placed in the right column (Table 2). For the font size task, the font type is Verdana. For the font type task, the font size is 12 pt, and the pictures task uses a 12 pt Verdana font.

To further elaborate on the story and picture content for task 3 on pictures, the story content is on changes in the Earth caused by the 2004 Asian tsunami. In the *on task* condition, the two pictures include (1) an aerial shot of damaged coastline, and (2) a heat map showing depth changes in the ocean floor. In the *advert* condition, we tried to select ads that were not obnoxious and reasonably consistent with the science and news theme. The ads were for National Geographic and the New York Times.

Table 2. Page layouts for font size and picture tasks.



2.2 Data Collection

Subjects participating in the experiment were employees of a major computer company, so they were computer literate and knew how to navigate the web. We recruited subjects at two of the company cafeterias (Fig. 2), offering them a cafeteria voucher in return for their participation. Our cafeteria location gave us a good cross section of company employees.

Fig. 2. Our gaze tracking setup at one of the company cafeterias.

The number of subjects, *N*, for each task is 82 (see Table 1). Participants did not receive all tasks, so the total pool of subject recordings was 132, 114 of which had usable eye tracking data (eye tracking quality varies by participant). Of these 114 participants, we had a good distribution of ages and gender: 74 male, 40 female, 2 subjects in their 20s, 63 in their 30s, 28 in their 40s, 18 in their 50s, and 3 above 60. Only a few had prior experience with eye gaze tracking.

Upon agreeing to participate, subjects were asked to sit at the eye tracking station shown on the right in Fig. 2. Subjects sit at a distance of around 60-70 cm from the monitor/eye tracker. The hood surrounding this station serves to (1) block out sunlight from interfering with the eye tracker, and (2) reduces distraction from the surrounding dining room. The experimental system then:

- Calibrates the eye tracker for the subject.
- Shows general instructions for the experiment: please read the stories, and expect a brief test of retention after each story.
- Presents a pre-test questionnaire asking for: name, first language, and a self-estimate of web usage.
- Presents a random subset of tasks to the subject. Each task consists of a lead-in page, the main testing page itself, and a post test of retention consisting of three multiple choice questions.

Our eye tracking setup includes the Tobii 1750 eye tracker [18], a camcorder taking a head-and-shoulders shot of the subjects, and three IBM T41 laptop computers. One laptop, the *gaze server*, was devoted to running the experimenter's console and the Tobii eye tracking software. A second machine recorded the camcorder video of the subject directly to hard disk in a compressed format. Finally, the third laptop, the *user machine*, was the machine used by our subjects. This machine ran an instrumented Internet Explorer browser, which recorded all URLs visited, HTML content, and dynamic events such as scrolling. WGA also recorded an event-driven movie recording of the user's screen through an adapted version of the VNC remote desktop system.

3 Results

Eye tracking information can be used to produce a detailed analysis of how the stories are read by the subjects. The first step is to determine when and what the subjects are reading. The WebGazeAnalyzer system includes reading analysis software that largely automates this process.

As mentioned in the introduction, eye movements during reading follow a characteristic pattern of fixations and saccades. When reading a line of text, a subject will fixate on the first word and then jump forward in the text around 8-12 characters to the next fixation point. Successive fixations will march down the line of text to the end of the line. Our analysis system finds this characteristic pattern of fixations by looking for a linear, horizontal grouping of fixations, calling the result a *gaze line*. Next, the analysis system uses a robust line matching algorithm to match gaze lines against lines of text from the story itself. From this reading detection and matching system, one can tell what was read, what was skipped, and how fast the subject was reading. In addition, we can look for:

- Regressions. These are backward motions in the text, and are indicative of confusion by the reader.
- Return sweeps. This is the motion from the end of one line to the beginning of the next. Return sweeps are known to be difficult on the reader for wide-formatted paragraphs.
- Distractions by pictures. Is the subject distracted mid-paragraph by pictures or other non-text elements on the page?

WGA uses two methods for measuring reading speed:

1. *Instantaneous speed*: a low-level measure of eye speed measured over fragments of gaze lines called forward reads (see [16,17] for details).
2. *Overall speed*: for a larger grouping of text (i.e. paragraph), the overall speed is the total text read at least once divided by the total reading time.

If a subject re-reads material, the overall speed suffers, because the amount of read text is not increasing, yet the total reading time is. The instantaneous metric, on the other hand, would record separate speed estimates for both the original read and the re-read (as they would have separate forward reads).

3.1 Font Size

Table 3 shows eye gaze reading statistics (plus retention) for task 1 on font size. In both the instantaneous and overall speeds, there is a slight trend for reading the larger fonts faster, but this trend is not significant. For example, the 14 pt font is read 8.5% faster than the 10 pt font, but this is not significant, $F(1,53) = 1.691$, $p < 0.2$. We found this lack of influence of font size on reading speed surprising itself, as we were expecting the larger fonts to be easier and faster to read. Furthermore, the regression rate, fraction of the material read, and retention in the post test are very similar across the font size conditions.

Table 3. Reading statistics under changes in font size. The larger fonts lead to a longer line length, which increases the time spent in return sweeps. Row measures with significant differences are shaded, and standard deviations are in parentheses.

Reading Statistic	Font Size			Significant difference?
	Small	Medium	Large	
Instantaneous speed (char/sec)	38.5 (7.9)	40.8 (11)	41.8 (10)	no
Overall speed (char/sec)	23.2 (6.1)	23.6 (6.8)	24.5 (8.6)	no
Regression rate (reg/sec)	0.39 (.20)	0.40 (.19)	0.38 (.16)	no
Total sweep time (sec)	3.48 (1.3)	4.22 (1.2)	4.66 (1.1)	yes
Fraction read (%)	94.8 (7.0)	93.7 (7.5)	93.1 (11)	no
Ave saccade length (char)	11.2 (2.4)	10.7 (2.8)	10.5 (2.9)	no
Retention (% correct)	89.2 (16)	90.1 (18)	88.9 (16)	no

The only significant difference is in return sweeps. Readers given the 14 pt font spend 34% more time in return sweeps than those given 10 pt, $F(1,53) = 12.16$, $p < 0.01$. In order to maintain line formatting across conditions, line length gets longer for the larger fonts, making these return sweeps more difficult for the reader. Beymer, *et al* [17] found this same effect when studying the effect of line length on reading. In Fig. 3 (left), we show a plot of sweep time vs. font size, and on the right we show histograms of the sweep times for the 10 pt and 14 pt conditions. In the peak near 50 ms, the eye is making a single saccade from the end of one line to the beginning of the next, while the peak at 200 ms involves an additional correction fixation and saccade [17]. The histogram for the 10 pt size is dominated by the single saccade near 50 ms, whereas the 14 pt size has a much larger fraction of sweeps

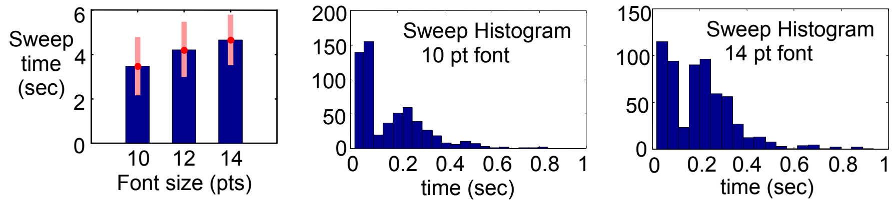


Fig. 3. The plot on the left shows that time spent in return sweeps increases significantly with font size (standard deviation is shown in light bars). On the right, histograms of return sweep times for the 10 pt and 14 pt conditions show that the larger font has a much higher proportion of return sweeps with a correction saccade and fixation – the peak near 200 ms.

clustered around 200 ms. In both the 10 pt and 14 pt condition, the angle α between consecutive lines (Bouma [19]) is about 2º (10 pt: 1.99º, 14 pt: 1.94º), so that is not causing the difference in sweep time.

Recall from the introduction the disagreement between Tinker and Rayner on the influence of font size on saccade length. Measured in terms of characters, Tinker claimed that it would get smaller (smaller perceptual span), and Rayner claimed that it would stay the same (saccade length scales with character size). In Table 3, we list average saccade length: 11.15 char for 10 pt and 10.49 char for 14 pt. While this is a 6% drop in terms of characters, this is against the background of a 45% increase in pixel size, so our data favors Rayner's interpretation. Also, there are no statistical differences between the saccade lengths across point sizes, as measured in characters.

3.2 Font Type

Using the overall speed metric, the serif font, Georgia, was read 8.45% faster than the san serif font, Helvetica, although this difference is not significant, $F(1,80) = 1.786$, $p < 0.2$. Likewise with the remaining statistics in Table 4, there are no statistically significant reading or retention differences between the two font types. Visually, the Georgia and Helvetica versions of the task 2 story are quite similar, so this result is not too surprising.

Table 4. Reading statistics under changes in font type. While there are no significant differences, the serif font was read 8.45% faster than the san serif font under the overall speed metric. Standard deviations are shown in parentheses.

Reading Statistic	Font Type		Significant difference?
	San Serif	Serif	
Instantaneous speed (char/sec)	38.7 (8.3)	40.9 (9.8)	no
Overall speed (char/sec)	20.1 (5.9)	21.8 (5.9)	no
Regression rate (reg/sec)	0.49 (.19)	0.45 (.23)	no
Total sweep time (sec)	3.52 (1.3)	3.71 (1.2)	no
Fraction read (%)	95.6 (3.9)	95.1 (6.8)	no
Ave saccade length (char)	11.3 (2.7)	11.8 (3.1)	no
Retention (% correct)	75.6 (16)	75.6 (15)	no

3.3 Pictures

How does the presence of pictures affect reading behavior? Table 5 shows some large significant differences between "on task" pictures and advertisements in terms of instantaneous speed and regression rate. Recall that the *on task* condition contains 2 pictures related to the text, positioned in the right column. In the *advert* condition, those 2 pictures are replaced by advertisements of the same size and aspect ratio. Under the advert condition, readers read 22% faster ($F(1,52) = 11.23$, $p < 0.002$) and have a 25% higher regression rate ($F(1,52) = 5.014$, $p < 0.03$) than the on task condition. Interestingly, in the overall speed, there is not a significant difference between the two conditions.

Table 5. For the pictures task, reading statistics reveal that advertisements speed up the reader's instantaneous speed and increase his regression rate. Row measures with significant differences are shaded, and standard deviations are in parentheses.

Reading Statistic	Pictures			Significant difference?
	On task	Advert	None	
Instantaneous speed (char/sec)	37.1 (6.3)	45.4 (11)	41.5 (10)	yes
Overall speed (char/sec)	21.2 (4.1)	23.0 (7.9)	22.5 (6.3)	no
Regression rate (reg/sec)	0.43 (.17)	0.54 (.19)	0.53 (.22)	yes
Total sweep time (sec)	5.29 (1.2)	5.64 (1.5)	5.35 (1.4)	no
Fraction read (%)	93.6 (7.8)	92.3 (11)	92.3 (12)	no
Retention (% correct)	80.2 (19)	80.2 (24)	80.9 (21)	no

How can we explain the speed difference between the advert and on task conditions? The on task reading speed is below average – 37.1 char/sec versus an average of 40.1 char/sec in tasks 1&2. Perhaps increased picture viewing is slowing the reading speed down. To measure picture viewing relative to reading, we use three metrics: (a) total time, (b) number of distinct picture visits, and (c) where in the text the reading interruption occurs. First, for the two pictures in the task 3 story, on task subjects spend 44% more time fixating on the pictures compared to advert subjects. Due to high variance in the time data, however, this difference is not significant (picture viewing is quite unstructured and variable among subjects). Second, subjects in the on task group have 60% more distinct visits to the pictures compared to the advert group, a significant difference ($F(1,52) = 4.471$, $p < 0.05$).

While the on task subjects spend more total time viewing pictures, we can also use eye tracking to measure where in the story the reader is interrupting themselves to view pictures. As shown in Fig. 4, we label reading interruptions relative to the story's paragraph and line structure, with the reasoning that some types of interruptions are more disruptive to reading than others. Viewing pictures before reading (B) or after reading (E) the text is not disruptive, while interrupting reading between lines (L) or inside a line (I) would arguably slow down reading. Fig. 4 (right) shows the distributions of picture interruptions; note that the on task condition is weighted more toward the (I, L, P) interruptions, which could help explain the slower average reading in the on task condition.

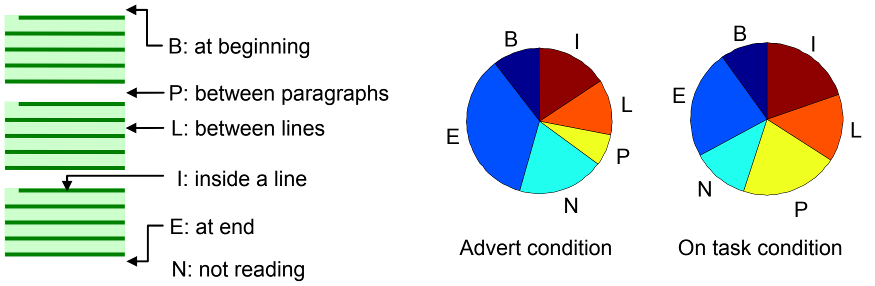


Fig. 4. Reading interruptions to look at pictures are categorized in 6 ways, depending on where the interruption occurs in the paragraph structure. On the right, we show the category distributions for the advert and on task conditions. Notice how the on task condition is more heavily weighted towards the more disruptive I, L, and P interruptions.

Turning to the subjects in the advert condition, the presence of ads increased their instantaneous reading speed – their average speed of 45.4 char/sec was higher than the average task 1&2 speed of 40.1 char/sec. This speed increase may be related to their increased regression rate. Viewing ads is more cognitively disruptive than viewing the on task pictures. This disruption, in turn, may exhibit itself as more regressions, as regressions are symptomatic of confusion by the reader. Indeed, we measured 20% more re-reading of material in the advert condition compared to the on task condition. We hypothesize that ads, a cognitive disruption, increases instantaneous reading speed through re-reading. On task pictures, on the other hand, do not increase regressions, and they slow the eye from a mechanistic standpoint, not cognitively.

In the discussion section, we explore the relationship between speed and regression rate in more detail.

3.4 Reading Statistic by Demographic

While our data analysis has so far focused on the experimental variables of font size, font type, and pictures, we also collected demographic data from the subjects. In this section, we look at reading behavior across age groups, gender, and the subject's first language. When investigating a particular demographic, we include reading data from all three tasks.

Table 6. Reading statistics for the different age groups. There are significant differences, with the older group reading faster, spending less time, and re-reading less. Row measures with significant differences are shaded, and standard deviations are in parentheses.

Reading Statistic	Age			Significant difference?
	30s	40s	50s	
Time (sec)	79.1 (25)	75.1 (24)	63.4 (28)	yes
Overall speed (char/sec)	21.1 (5.5)	22.4 (6.4)	26.9 (8.9)	yes
Regression rate (reg/sec)	0.46 (.18)	0.44 (.20)	0.38 (.15)	no
Fraction re-read (%)	36.4 (13)	32.6 (13)	24.5 (12)	yes
Retention (% correct)	82.9 (14)	80.5 (14)	83.0 (7.5)	no

When analyzing reading by age, we focused on the age groups of 30s, 40s, and 50s, as the tail ends of the distributions (20s, 60s) had small sample sizes (2 and 3, respectively). Table 6 shows reading statistics for the age groups using the reading data from all tasks. While retention is flat across age, there is a linear trend for all of the remaining statistics, and there are a number of significant differences between the 30s and 50s groups. Subjects in the 50s group read faster than the 30s group and they spend less total time on the story (for overall speed, $F(1,79) = 11.03$, $p < 0.002$). Subjects in the 50s group re-read significantly less than the 30s group, $F(1,79) = 11.42$, $p < 0.002$. While the 30s group has a 48% higher regression rate, this difference is not significant due to high variance. Overall, the 50s group performs the highest, matching the retention scores of the other groups but using less reading time.

Analyzing the reading data by gender (74 male, 40 female) reveals practically no differences. The only significant difference is in the total story sweeptime, with males having an average of 4.2 sec and females an average of 4.7 sec ($F(1,112) = 5.413$, $p < 0.03$). For some reason, males have an easier time repositioning themselves at the beginning of the next line in a return sweep. There are no differences in retention or speed.

To analyze online reading by first language, we divided the subject pool into two groups: (1) those reporting English as their first language (67 subjects), and (2) those reporting some other language (47 subjects). Table 7 shows a number of significant differences between the two groups, with the English first group reading faster, spending less total time, and re-reading less than the non-English first group. (For example, the overall speed of the English first group is 29% faster than the non-English first group, a significant difference, $F(1,112) = 23.6$, $p < 0.000005$.) There is no difference in retention or regressions, so the non-English first subjects simply have a time handicap compared to the English first group.

Table 7. Reading statistics broken down by whether English is the subject's first language. Subjects whose first language is English read faster and did less re-reading. Regression rates and retention are the same for the two groups. Row measures with significant differences are shaded, and standard deviations are in parentheses.

Reading Statistic	Is English the subject's first language?		Significant difference?
	yes	no	
Time (sec)	65.4 (19)	90.6 (27)	yes
Instantaneous speed (char/sec)	43.5 (9.8)	36.4 (8.2)	yes
Overall speed (char/sec)	24.6 (6.2)	19.0 (5.7)	yes
Regression rate (reg/sec)	0.45 (.20)	0.45 (.18)	no
Fraction re-read (%)	30.7 (13)	38.5 (14)	yes
Retention (% correct)	82.7 (11)	82.2 (15)	no

4 Discussion

In this paper, we investigated how the typographical issues of font size and font type impact online reading. Using eye tracking, we measured reading speed and other eye gaze statistics, getting a detailed look at the reading process. We also studied how different types of pictures – ads and relevant pictures – affect reading and retention.

In our study of font size, our results confirm Morrison and Rayner's claim that saccade lengths should scale with the point size. Also, the lack of a significant difference in speed across font sizes may tempt a designer to use a smaller font in order to cram material on one page. However, while we did not quantitatively estimate subject font size preference, the reaction to the small 10 pt font was fairly negative. This is especially true since eye tracker constrains the subject from getting closer to the monitor. We feel that a combined measure of speed and preference would probably select the 12 pt font.

For font type, our study of serif vs. san serif yielded no significant differences in the eye tracking or retention metrics. There was an 8.45% advantage in overall speed for the serif font, but the difference was not significant; perhaps using $N >> 82$ would increase the sensitivity of the test.

When studying the distracting influence of pictures on reading, we found that pictures related to the text slowed down instantaneous reading speed but did not increase regressions. While these pictures caught the reader's eye and slowed them down, they were not a cognitive load. Advertisements, on the other hand, increased regressions and caused more re-reading. We believe these factors, in turn, increased instantaneous reading speed. While this is indicative of an increased cognitive load, there was no retention difference between the two conditions.

This analysis of the picture task raises a more general question: what is the relationship between reading speed and the regression rate? To investigate, we used the reading data for all tasks, and we divided our entire subject pool into two groups, *reg-low* and *reg-high*, depending on whether the subject's average regression rate was below or above the overall median rate of 0.44 reg/sec. There is an interesting reversal when comparing *reg-low* and *reg-high*'s speed measures. The *reg-low* group had a significantly lower instantaneous speed than *reg-high*, $F(1,112) = 7.32$, $p < 0.01$. However, the *reg-low* group had an overall speed that was 10.7% higher than *reg-high*, a nearly significant difference, $F(1,112) = 3.36$, $p = 0.069$. Also, *reg-high* re-read a significantly higher fraction of the material compared to *reg-low*. Overall, this says that regressions slow the overall reading speed down, but the reader increases their instantaneous speed as they are re-reading the material.

Returning to the issue of distraction from advertisements and banner blindness, it is interesting to note the rise of contextual advertising and its potential to increase the attraction of ads. In contextual ads, the web page content is scanned to determine those ads that may interest the reader, and only those ads are presented (Google AdSense, Chitika [20]). Thus, the ad will target the web page's intended audience, potentially creating a distraction that is hard to resist. The effect of contextual ads on a subject's performance and task completion would make an interesting future eye tracking study.

References

1. Tinker, M. Legibility of Print. Iowa State University Press, Ames, Iowa (1963)

2. Paterson, D.G., Tinker, M.A. Studies of Typographical Factors Influencing Speed of Reading: II. Size of Type. *Journal of Applied Psychology*, 13:2 (1929), 120-130.
3. Legge, G.E., Pelli, D.G., Rubin, G.S., Schleske, M.M. Psychophysics of Reading: I. Normal Vision, *Vision Research*, 25:2 (1985), 239-252.
4. Bernard, M., Lida, B., Riley, S., Hackler, T., Janzen, K. Comparison of Popular Online Fonts: Which Size and Type is Best? *Usability News*, 4.1, Software Usability Research Laboratory, Wichita State University (2002)
5. Paterson, D.G., Tinker, M.A. Studies of Typographical Factors Influencing Speed of Reading: X. Style of Type Face. *Journal of Applied Psychology*, 16:6 (1932), 605-613.
6. Boyarski, D., Neuwirth, C., Forlizzi, J., Regli, S.H. A Study of Fonts Designed for Screen Display. *Proc. CHI*, Los Angeles (1998), 87-94.
7. Research-Based Web Design and Usability Guidelines: Chapter 14. Graphics, Images, and Multimedia, U.S. Department of Health and Human Services, http://www.usability.gov/pdfs/guidelines.html
8. Benway, J.P. Banner Blindness: The Irony of Attention Grabbing on the World Wide Web. *Proc. Of the Human Factors and Ergonomics Society 42nd Annual Meeting*, vol. 1 (1998), 463-467.
9. Burke, M., Hornof, A., Nilsen, E., Gorman, N. High-Cost Banner Blindness: Ads Increase Perceived Workload, Hinder Visual Search, and Are Forgotten. *ACM Trans. On Computer-Human Interaction*, 12:4 (2005), 423-445.
10. Rayner, K. and Pollatsek, A. The Psychology of Reading. Lawrence Erlbaum Associates, Hillsdale, NJ (1989)
11. Paterson, D.G., Tinker, M.A. The Effect of Typography upon the Perceptual Span in Reading. *American Journal of Psychology*, 60 (1947), 388-396.
12. Tinker, M.A., Paterson, D.G. The Effect of Typographical Variations upon Eye Movement in Reading. *Journal of Educational Research*, 49 (1955), 171-184.
13. Morrison, R.E., Rayner, K. Saccade Size in Reading Depends upon Character Spaces and Not Visual Angle. *Perception and Psychophysics*, 30:4 (1981), 395-396.
14. Eyetrack III: Online News Consumer Behavior in the Age of Multimedia. The Poynter Institute, http://www.poynterextra.org/eyetrack2005/main.htm
15. Norman Nielsen Group, http://www.nngroup.com
16. Beymer, D. and Russell, D. WebGazeAnalyzer: A System for Capturing and Analyzing Web Reading Behavior Using Eye Gaze. In *CHI 2005 Extended Abstracts*, ACM Press (2005), 1913-1916.
17. Beymer, D., Russell, D., Orton, P. Wide vs. Narrow Paragraphs: An Eye Tracking Analysis. *Proc. Interact 2005 – International Conference on Human-Computer Interaction*, Rome (2005) 741-752.
18. Tobii 1750 Eye-Tracker, Tobii Technology, http://www.tobii.com
19. Bouma, H. Visual Reading Processes and the Quality of Text Displays. In *Ergonomic Aspects of Visual Display Terminals*, E. Grandjean and E. Vigliani, Eds. (1980), 101-114.
20. Tedeschi, B. Internet Banner Ads Look to Get More Interesting (and Thus Less Easy to Ignore). *The New York Times*, July 11, 2005.